Filing Receipt

Received - 2021-08-16 04:29:57 PM
Control Number - 52373
ItemNumber - 59

PROJECT NO. 52373

REVIEW OF WHOLESALE ELECTRIC MARKET DESIGN	**§ §**	**PUBLIC UTILITY COMMISSION OF TEXAS**

COMMENTS OF ABLE GRID ENERGY SOLUTIONS

COMES NOW Able Grid Energy Solutions (Able Grid) and files these Comments in response to the Commission's Questions for Comment filed in this proceeding on August 2, 2021.

Executive Summary

Able Grid believes that ERCOT currently has almost all the necessary tools to improve reliability. Dramatic reliability improvements can be achieved by adjusting market volumes & dispatch strategy as well as accelerating the deployment of grid improvements like ECRS and Real-time Co-Optimization. Able Grid summary of recommendations:

- Cap the deployment of RRS resources at 30 minutes – 1hr max.
- Accelerate implementation of NPRR 1002 and consider the acceptance of ESR charging at the direction of ERCOT, for the sake of improving RT PRC even during EEA3.
- Accelerate the Implementation of ECRS.
- Accelerate the implementation of Real Time Co-Optimization
- Increase ECRS and non-spin volumes consistent with the responsibility of recovering RRS deployed MW.
- To the extent new reserve products are needed, apply non-discriminatory least-cost frameworks to ensure the market can procure the most efficient solution.

Introduction

Able Grid Energy Solutions (Able Grid) is a large-scale stand-alone energy storage developer active in all US wholesale markets. Able Grid develops transmission connect energy storage resources to serve as flexible capacity which improve system reliability.

Comments

1. **What specific changes, if any, should be made to the Operating Reserve Demand Curve (ORDC) to drive investment in existing and new dispatchable generation? Please consider ORDC applying only to generators who commit in the day-ahead market (DAM). Should that amount of ORDC-based dispatchability be adjusted to specific seasonal reliability needs?**

Increasing minimum operating reserves and products which contribute to Physical Responsive Capability (PRC) volumes during seasons of high Intermittent Renewable Resources (IRR) penetrations may be warranted to ensure a liquid real-time (RT) market when faced with uncertain supply. However, adjustments to ORDC in this respect ought to follow the currently implemented, non-discriminatory, framework for ORDC seasonal probability distribution adjustment. This current process is rational, fair, and transparent for all market participants.

The purpose of the Operating Reserve Demand Curve is to incentivize RT availability of reserves to mitigate possibility of load shed. The price of ORDC represents the probability of loss of load (LOLP) multiplied by the "cost" of unserved load which is determined to be $9,000/MWh. The probability distribution used to determine this LOLP is constructed using historical ERCOT reserve forecast error data broken into seasons. Reserve forecast error is considered the difference between what reserves are expected on an hour ahead basis (HA reserves) and what reserve quantities actual exist in RT operations. Historical data of reserve error used to determine the seasonal distributions dates to 2010 and is updated after each new season. To the extent that periods of larger IRR penetrations contribute to greater uncertainty of reserves, this is already being considered in today's market practices of ORDC.

Given the focus on reliability in today's marketplace, ERCOT will likely require a more flexible ORDC probability distribution framework to accommodate the rapidly evolving generation mix. This is readily achievable by changing the empirical dataset which we currently use to feed the seasonal probability distributions. Today we use a comprehensive dataset dating back to 2010. If we used a rolling dataset (3yrs of trailing historical data as an example), then the new reserve forecast data updated each new year will have a greater influence on adjusting the ORDC probability distributions. If reserve forecast error is increasing, this will result in ORDC prices kicking in at higher levels of RT reserve volumes, thus incentivizing the RT availability of more resources to SCED.

ORDC should ***not*** be limited to certain types of resources. Removing ORDC value from resources generating in real-time would ultimately create systemic problems for billions of dollars of hedging arrangements currently in the market today. Any action which reduces the liquidity of the financial markets surrounding operation of the ERCOT grid would cause significant damage to the Texas investment environment. Additionally, ORDC opportunity provides an important price signal to non-dispatchable resources as well. As wind penetration grows in the ERCOT marketplace, periods of RT reserve scarcity are becoming increasingly correlated with low RT output from wind resources. Two important take-aways from this trend:

i. The design of ORDC negates the need to limit compensation to specific resources. Resources which don't show up during periods of scarcity to not enjoy the benefits of ORDC. Additionally, if the wind resources agreed to any DA sale of energy, they would be required to buy that energy back at the RT price including the ORDC. This creates a very strong incentive for wind resources to bid into the DAM with as much certainty as possible.

ii. ORDC opportunity creates an incentive to site IRR resources in locations which will give them better incidental availability during periods of scarcity. By way of example, assume that summer months are expected to be tight on reserves. West Texas wind resources generally perform their worst during the summer months and the lack of west Texas wind generation during peaking demand can help cause ORDC pricing (Summer 2019). As a wind developer, this creates an incentive to site and develop projects in other geographies which have a more complementary generation profile with periods of expected scarcity. Southern and Coastal wind patterns have a higher correlation with periods of scarcity and therefore the market has a clear signal to favor generation development in these regions rather than the West. Without exposure to ORDC, this would not be a part of the decision-making process for developers.

    This same logic applies to the siting of solar as well, east to west siting of solar resources ought to consider how the generation profile overlaps with the possibility of scarcity in the marketplace. Resources sited in the west would enjoy a generation profile which better serves the afternoon peak demand of ERCOT. It's worth noting that these locational ORDC benefits would obviously be evaluated against all other risks an IRR projects wears like transmission availability. Regardless, it is still critically important that ORDC value is considered in their investment decisions so that the market actors are ultimately moving forward with the most "valuable" projects when considering all market signals and risks.

As implemented today, capacity which is committed to financial obligations in the DAM is explicitly ***not*** entitled to ORDC in the real time. The expectation of ORDC in the real-time informs participants bids and offers in the DA which ultimately leads to higher prices in the DA during days of expected scarcity. By taking on a DA obligation to serve energy or Ancillary Services during real time operations, the resource is accepting the DA prices as incentive to be online and operating during the Real-time. Attempting to make ORDC value exclusive to these DAM resources would be double compensating resources which already committed to be available.

2. Should ERCOT require all generation resources to offer a minimum commitment in the day-ahead market as a precondition for participating in the energy market?

a. If so, how should that minimum commitment be determined?

b. How should that commitment be enforced?

The DA market is currently an opportunity, not an obligation, to get economic certainty for operating in the Real-Time. Obligating resources to commit capacity in the DA, beyond what they would commit under their own rational, could consequentially reduce ERCOTs real-time awareness of available capacity. Resources which cannot guarantee availability in the real-time should not commit in the DA because then the supply/demand equilibrium in the DA is skewed by the "phantom supply" (phantom supply in this context refers to resource which commits capacity in the DA because of a market rule requirement rather than actual certainty that they can be available during Real-Time operations). If these uncertain resources don't participate in the DAM, then the DA demand for Energy and AS can be cleared and served by supply resources which can commit to DA obligations under their own intentions.

Ensuring all DAM participation is voluntary also maximizes the available capacity in the Real-Time. By way of example, if the operating day is expected to be a very low wind day, Wind resources will be very hesitant to bid into the DAM. By reducing the supply available to clear the DAM, this presents dispatchable resources with the opportunity to clear their capacity at prices which makes them economically whole. These DAM incentives create available dispatch capacity in the RT to manage the uncertainty of wind generation. If the wind does show up in the real time, the market enjoys low RT prices and capacity resources can "buy-back" their DA obligation at a profit. If wind generation doesn't show up at all in the RT, the capacity which was committed in the DAM is ready and available to serve demand. In the alternative circumstance, where wind resources were required to commit capacity in the DAM, the grid would be relying solely on the liquidity of the RT market to ensure that capacity shortfalls are met. Dispatchable resources, which didn't clear the DAM due to forced commitment from wind capacity, would be relied upon in the

RTM even though they never received a DAM payment to incentivize their availability. This alternative market procurement strategy is a much riskier operating strategy and would ultimately reduce the reliability of the marketplace.

The operating strategy proposed in the initial question would create the incentive for each individual market resource to procure their own guaranteed capacity to serve their DAM obligations. This would be grossly inefficient from the perspective of optimizing the utilization of available market resources. Procurement of capacity via an organized Day Ahead Market provides a centralized expectation of services provided by capacity resources. Some capacity is utilized to serve short term needs (PFR, Regulation, Peaking Capacity), while other capacity resources serve longer term needs (clearing on-peak DAM energy needs). Leaving market participants with the responsibility to procure their own firm capacity will disincentivize the continued growth of cost-effective "energy only" resources and lead to extra costs incurred from thinking about capacity as service required by each individual resource rather than a capability which must be met by the portfolio at large.

3. What new ancillary service products or reliability services or changes to existing ancillary service products or reliability services should be developed or made to ensure reliability under a variety of extreme conditions? Please articulate specific standards of reliability along with any suggested AS products. How should the costs of these new ancillary services be allocated.

ERCOT reserve scarcity and ORDC accounting is based off the concept of always ensuring adequate operating reserve and Physical Responsive Capability (PRC) volumes. PRC is critically important to the operations of the grid because even during the tightest market supply /demand conditions, reserves capable of arresting frequency decay must remain available to address possible contingencies. Once PRC volumes drop below certain thresholds, it is prudent for the grid to enact load shedding measures rather than let PRC drop further. If the market were to completely deplete its PRC volumes, and a contingency on the system were to occur, there would be nothing to stop the decay of frequency and the system would rapidly devolve into uncontrolled blackouts and catastrophic failure. This is the dire scenarios which ERCOT narrowly avoided during the early hours of Feb 15th, when load shed was first implemented. Part of the reason that the market became so tight during these moments was because RRS resources, those which are procured to specifically provide frequency response reserve capacity (PRC MW), had already been deployed in the late hours of Feb 14th leading up to the critical moments at 1AM on the 15th. When these RRS resources were deployed, the instruction from ERCOT was to stay deployed for hours on end. The resources were not able to recall their capacity and make their response capability available again because the market was using these resources to serve energy needs rather than PFR/PRC.

Given that the PRC volumes are the critical measuring stick for determining load shed actions, the market should be equipped with strategies necessary to recall RRS resources as soon as possible and allow them to make the MW available again for future potential PRC scarcity as soon as possible. When we think about the supply stack of resources that can provide PRC, it is only a subset of resources which can respond in the quick manner needed to address frequency needs. From this perspective it is easy to see that this supply stack is at the greatest risk of scarcity when compared to the larger energy supply stack because the PRC supply stack is only a small subset of the energy supply.

In an ideal world, dispatch of follow-on AS products such as ECRS and non-spin would be deployed to recover the RRS resources as soon as possible. After the first brief period of deployment, RRS resources are no long providing a unique service to the market. Beyond the 5 – 15 minute time frame, SCED should be able to dispatch slower ramping resources to meet the "energy need". Slow responding, non-PRC-accredited AS like non-spin resources should take over the energy responsibility that the RRS resources are serving in the moment, allowing the market to recover the "scarce" MW which are the rapid responding RRS MW which are PRC accredited. The replacement of the RRS deployment in this circumstance would *increase* the amount of PRC available to the market in the real time.

Regardless of price implications for the market, having more PRC available during the moments of 1:30AM would've allowed the ERCOT staff and participants to act with less haste because the market still has the critical buffer needed to respond to sudden frequency shocks. By way of example, consider the scenario where 100MW of RRS is deployed to meet frequency needs. PRC was 100MW before the deployment and now PRC is at 0 because the 100MW RRS resource is providing energy with no upward capacity left. It would be imperative for the market to recover PRC capability to avoid load shedding and so therefore SCED should dispatch replacement non-spin MW to take over the energy serving needs of the market so the RRS MW can make themselves available again (important to note that non-spin capacity is not counted as PRC because it takes 30 minutes for non-spin to become available). Once 100MW of non-spin is online and generating, the 100MW RRS resource can back down to 0 output and the PRC available to the market increases to 100MW.

It should be noted that Energy Storage Resources (ESRs) are the best resources available to the market for providing frequency responsive capacity. The response time of *battery* energy storage is an order of magnitude faster than the response time possible from the next fastest resources (hydro operating in sync-condenser mode). The issue with energy storage participation during February was the inability to charge storage resources in an efficient manner due to market rules and the prolonged deployment of RRS experienced in the hours leading up to load shed.

It is critical to stress the value of ESRs, even when charging. The purpose of a battery charging during such tight conditions is to improve the resources availability over the course of the event. Even during the charging process, the ESR is providing reserve capacity to the market via demand response capability. The ESR can halt charging at moment's notice to provide near-instantaneous frequency responsive capacity. Deploying additional non-spin for the sake of charging ESRs reduces the available supply stack left but leaves the market in the exact same position with regards to available reserve/frequency responsive capacity. Once the ESR is charged again, the market is in an improved capacity position given that the generation used to charge the battery is now available to serve load or reserves. Not only that, but the battery is now able to offer discharge capacity once again.

Assuming that an ESR is given the opportunity to charge before their storage energy is completely depleted, a charging ESR can provide 2x its nameplate in PRC capacity. This is because a charging ESR can act as demand response and halt charging at moment's notice to address frequency. From there, the ESR can then go from 0 output to full discharge. This flexibility creates the ability to seamlessly transition from full charge to full discharge creating 2x nameplate in frequency responsive capacity. By way of example again, consider a 100MW 2hr energy storage resource providing RRS. The 100MW storage resource gets deployed for 100MW of RRS and PRC goes from 100 to 0. Non-spin resources should be dispatched to the market at the same time as RRS to ensure available MW to replace the RRS deployed MW in 30 minutes. Even if the RRS deployment lasted an hour, a 2hr ESR would still have 1hr of charge left in the battery after recovering from the RRS deployment. This means that the ESR resource is immediately available to provide another 100MW of RRS deployment once recalled. In addition, it would be prudent market operation to replenish the state of charge for energy storage resources so that they can maximize their availability for the marketplace going forward.

As a general rule, tight conditions which are *intentionally* created by the market (by allowing ESRs to charge during EEA conditions) for the sake of increasing capacity available to the market for future use, is better than uncontrolled tight conditions experienced due to lack of available capacity. From that perspective, the market would need 200MW of non-spin to replace the deployment of the 100MW ESR, plus the subsequent 100MW of ESR charging. In this moment, by allowing 100MW of ESR charging, 200MW of PRC is created by the ESR. In the described dispatch strategy, PRC is maintained in an efficient manner and non-spin is procured and deployed to meet netload forecast error and recovery of RRS resources. This methodology would seem to improve the PRC availability to the market especially during scarcity conditions. With this strategy in mind, additional procurement volumes for non-spin could be calculated based on RRS volumes and % of RRS volumes being provided by ESRs.

Accelerated implementation of ERCOT Contingency Reserve Service (ECRS) would also improve the dispatch and availability of critical reserves during periods of scarcity. ECRS is expected to provide 10-minute ramping capacity to help meet ramping uncertainty and to recover RRS resources. Prioritizing the implementation of this new AS would aid the amount of ramping capacity available to SCED and would relax the inefficient use of RRS for prolonged energy related deployments. RRS resources should solely be used to address frequency.

The Value of Lost Load (VOLL) in ERCOT is assigned at $9,000/MWh. It is reasonable to say that value of losing reserves used to maintain system frequency are more costly than the Value of Lost load. This is because even in the face of load shedding, the market must maintain a minimum number of reserves available to arrest frequency during a contingency. If the market were to completely deplete its PRC volumes, and a contingency on the system were to occur, there would be nothing to stop the decay of frequency and the system would rapidly devolve into uncontrolled blackouts and catastrophic failure which ERCOT itself has said could last an undetermined amount of time (true black start has never been attempted in modern ERCOT history). Therefore, it is justified to charge ESRs even during period of rotating blackouts. Recharging ESRs provides PRC capacity while charging, and at the same time creates more capacity for future use and therefore a greater load carrying capability in the face of prolonged extreme scarcity.

As a fundamental principle of retail competition market design, costs should be allocated to the end users of the wholesale power system. LSEs (Load Serving Entities) are ultimately the parties responsible for procuring supply and delivery of electricity on behalf of their customers on a 24/7 basis. The goal of the LSE is to minimize the total cost of serving their customers' electricity needs through strategies which combine long term offtake agreements with market purchases to meet variable demand on a moment-to-moment basis. Even without additional AS products, LSEs are already exposed to the cost implications of extreme conditions. When customer demand exceeds expectations, LSEs are required to make market purchases. The LSE should have the opportunity to decide how to procure its mix of energy + firm delivery in the most cost effective and reliable manner. One LSE may elect to procure energy and firming capability from a single resource like a CCGT. On the other hand, an LSE might determine that the most economic and reliable solutions is to procure the cheapest possible MWh from renewable resources while procuring firming services from a battery resource or a peaking gas plant. It's imperative for a competitive retail market that the end users face these costs directly for the sake of efficient and transparent risk management.

To the extent that ERCOT introduces new AS products, the additional MW will be used as reserves to meet unexpected shortfalls in supply. Given that LSEs already wear the cost of meeting demand when it exceeds expectations, they ought to be allocated this new AS cost of ensuring ample supply of reserves in the market to meet uncertainty. LSEs faced costs of $50B from the Feb cold weather event. The justification for new AS products would be to prevent future events of extreme scarcity from occurring again so costs associated with new reserves would essentially replace the current scarcity costs that ERCOT LSEs are already facing. AS product costs would theoretically be much more consistent and manageable than the scarcity prices experienced recently in ERCOT. By directly allocating costs to the LSEs, it also provides the end users/LSEs with the best opportunity to manage and hedge risk consistent with each companies' appetite for risk. Allocating the costs of new AS products to any market participant besides the end use customer hinders the ability of market buyers to manage the risks they face. Allocation of AS costs to generation resources distorts the true cost of energy provided from those resources and robs the LSEs of the ability to actively manage the cost of AS under their own strategies.

6. How can the current market design be altered (e.g., by implementing new products) to provide tools to improve the ability to manage inertia, voltage support, or frequency?

It's important to note that the concept of "inertia" as a critical grid service is somewhat of a misnomer. It's not that inertia is a service which is fundamentally required to maintain a synchronous AC power system. Rather, the inertia from generator rotating masses on the power system provide a valuable contribution to slowing down the pace of initial frequency decay when a contingency occurs because those rotating masses set the primary frequency for the system. This is essentially a Primary Frequency Response (PFR) service which could be replaced with other PFR-capable resources during periods of low inertia. With the growth of "grid-forming" inverters and quick responding energy storage resources, the market is increasingly equipped to serve frequency needs without rotating masses on the system. Fast Frequency Response (FFR) as a service is uniquely positioned to offset the decay of inertia in the market. Studies presented at the ERCOT PDCWG suggest that increasing FFR participation can significantly increase the grids tolerance for low inertia and possibly even reduce required RRS volumes.

There is nothing inherently right or wrong about serving PFR needs through generator rotating mass inertia or by other means, the important point is that the market is equipped with the real time awareness of system needs and can address those needs through market mechanisms to ensure the lowest cost solution. Energy Storage Resources can respond to system needs faster than any other resource available in the market. Services like the FFR (Fast Frequency Response) carveout of RRS (Responsive Reserve Service) can aid in reducing the need for inertia on the system. Procuring capacity which can respond to frequency excursions on the millisecond basis provides a "synthetic or virtual" inertia which slows down the Initial rate of change of frequency the same way actual inertia does. The faster a resource can respond to native changes in frequency, the more significant contribution they make to addressing frequency stability needs. ERCOT and the Commission should implement a market mechanism which incentivizes all resources capable of providing this type of PFR service to be available to the system at all times based on hour of the day and season.

The issues of voltage and frequency support on the system also tie back to the response to question 3 above, advocating for ample supply and efficient dispatch of reserve assets. Maintaining stiff voltage sources and quick response capability for frequency changes is ERCOT's first (and arguably most important) line of defense against blackouts and an unreliable power system. Even at the point of firm load shed, ERCOT must retain a minimum level of PRC (Physical Responsive Capability, reserves which can respond to frequency). During the early morning of Feb 15th when ERCOT started to enact load shed, capacity available to arrest frequency decay would've provided a critical buffer for control room decision making. Even if this frequency support was only available for a brief period, a stronger grid frequency would've prevented tripping of generators which were damaged by the rapid swings in frequency and may have allowed for a more measured approach to load shedding strategy.

Conclusion

Able Grid appreciates the opportunity to provide these Comments and looks forward to working with the Commission and other interested parties on these issues.

Respectfully submitted,

Samuel Brandin
Manager of Strategy & Analytics
Able Grid Energy Solutions
SBrandin@ablegridenergy.com
(720) 675-3800